北星短大生の英語習熟度

── 能力テストからの考察 ──

清　瀬　　健

1. 序

1970年代初頭までの外国語能力テストの主流は，言語の要素（音素，抑揚，強勢，形態素，語，文，節，句，等）を個別にテストすることが強調されていた。又，聴く，話す，書く，の4技能もそれぞれ別々に測定されることが良しとされていた。これは部分的測定法 (Descrete-Point Approach) と呼ばれ，Lado らによって強く主張され，彼の Language Testing (1961) は，外国語教育にたずさわりテストを作成する者に多大の影響をあたえた。もちろん現在もこの考え方が多くのテストの形式に反映され，世界的規模で行なわれている TOEFL もこの考え方にもとづいて作成されている。部分的測定法の理論的基盤となるものは，高い信頼性 (reliability) と妥当性（この場合，内容的妥当性 content validity）である。それ以前のテストの方法は，テスト作成と採点に主観性が入り込みやすいタイプの翻訳や自由作文などであり，信頼性，妥当性，経済性などにあまり関心が払われずにただ慣行にしたがって行なわれてきた。

この部分的測定法に対し，1960年代後半に入ってから次々と批判的な論文が発表されるようになった。心理測定学―構造言語学の流れをくむ部分的測定法に対する批判は，言語の要素をバラバラにしてその個々のデータから外国語の能力全体を測定するには，あまりにも一面的すぎ，かならずしもその人の言語能力全体を測定している訳ではない，というのである。つまり言語の能力を測るためには，ある項目の知識を中心に測るのではなく，いくつもの知識を同時に駆使し，時間の制約のもとで機能させることのできる能力を測る integrative（統合的）なものでなければならない，という主張だ。

この統合的測定法テストとしてにわかに脚光を浴びたのが，書き取り (dictation) とクローズ (Cloze) テストである。Oller [Oller, 1971] は UCLA の英語クラスわけテスト (ESLPE) の語彙，文法，作文，音韻，書き取りテストの各部門の相関及び，この5部門の総計の相関を調べた結果，書き取りが，作文と総計の相関を除いて，すべての部門で一番高い相関を示したと報告している。又，Oller [Oller, Irvine, Atai, 1974] によれば，TOEFL とクローズテストの相関が報告されており，この中で聴き取り，構造，語彙，読解，作文能力の各部門と，その部門を除く TOEFL の合計の相関を調べ，この各部門を除く合計とクローズテスト（適語法）との相関を見たところ，聴き取りとそれを除く合計との相関以外はすべてにクローズテストが一番高い数字を示したことを報告している。この他にも Stubbs, Tucker [Stubbs, Tucker, 1974] らの多くの研究報告がなされ同様の結果がみられる。つまり書き取り，クローズテストともかなり総合的な外国語能力を測定しているということが明らかになってきている。

本稿は，このような観点に立って本学学生の英語能力テストを行ない，その結果を分析し，まとめたものである。

2. 目　　的

本稿の目的は次の3点である。

1）本学学生の総合的英語能力を測定するために能力テストを実施し，その分析を行う。

2）能力テスト受験者中における本学学生の相対的な英語能力を，公開されたデータから明らかにする。

3）分析結果から，今後のカリキュラムの方向を考える。

3. 試験実施にあたって

3.1 能力テストの条件

測定に用いるテストの選定は，序で述べたように，総合的な英語能力を測ることができると同時に，本稿の目的にかなうテスト，すなわち次の条件をみたすよう考慮した。

1）信頼性があり，妥当なもの
2）クローズテストや書き取りを含んでいるもの
3）さらに文法力，読解力，聴解力等の能力を個別に測定できること
4）データが個人はもとより，団体としても処理され利用できること
5）学習効果[1]による点数の増減を考慮しなくてよいもの
6）受験料が学生に大きな負担とならないこと

上記の条件にさらに補足的に説明を加えるならば，1）の信頼性とはテストの得点の安定度で，いつ何度受験しても，採点者が誰であれ，同様の結果が得られることである。また，妥当なテストとは，測定しようとする部分を正確に測定できるものであり，外部テストとの相関などで統計的に明らかにされ，そのデータが公開されているものである。さらに，問題の作成にあたっては，項目分析を行い，その結果として高い弁別力をもっていること。そして，本学学生の英語能力を正確に測定するために適度の困難度をもっているものがのぞましい。2）のクローズテストと書き取りは，序で述べたように，英語の総合能力を測定できると仮定した場合，文法力や聴解力といった個々の能力との比較が可能になるためである。したがって，できることなら，3）で上げたそれぞれの能力を測定できるものが好ましい。4）では，本学学生と他の受験生集団との比較を試みる際に，あらかじめ本学学生だけの統計的な処理がなされているならば，その比較が容易になると考えた。5）の学習効果とは，注1）で記した通りであるが，具体的には，市販されている問題集をやることによって生じる点数の増加のことである。したがって問題集などが手に入らない能力テストが適当であると考えた。6）では，経済的な理由から，できるだけ料金が安く，しかも他の条件をも満たしてくれるものである。

3.2 主な能力テスト

現在日本で広く行なわれている英語能力テストには，ETS (Educational Testing Service) が開発し，実施している TOEFL がある。TOEFL は日本ばかりでなく世界的規模で行なわれており，主にアメリカの大学で留学，研究をしに行く人々を対象に行なわれている。大学で授業を受けたり，研究生活をする上である程度不自由なく意志の疎通が行なえるかを判定するためのものであり，したがってテストは，レベルの高い人々を照準においていると言われており，本学学生にとっては多少困難すぎるように思われる。また，受験料でも負担が大きすぎる。

TOEIC (Test of English for International Communication) は，日本人のために ETS が開発したものであり，留学が目的でなく，単に英語能力を測定したいという人々にとっては，TOEFL より適当なテストと思われるが，TOEFL 同様，受験料の負担は学生にとって大きい。

実用英語検定（英検）は，他の能力テストに比べ日本での歴史が古く，文部省の認定ということもあり，受験者数も多い。1985年度は全国で152万人が受験している。1級から4級までの問題は毎年公表され，問題集となって市販もされている。英検の場合，試験の信頼性や妥当性を示す統計資料は，本稿で述べる他のテストのように公開されていない。

また，大学生を対象にしたものには，大学英語教育学会 (JACET) が開発し，毎年春秋2回行なわれている JACET-COLTD Listening Comprehension Test がある。1985年12月実施テストでは8,690人が受験した。データの処理は，国，公，私立の短大，4年制ごとの学年次でなされており，さらには専攻別に英語専攻，それ以外の文科系，理科系，その他と訳けて行

われている。しかしテストは名前が示す通り，聴解力のみの試験である。

一般には広く普及しているとは言えないが，Harris (1969: 13–23) が述べている適切なテスト[2]として条件をみたしているものに，YMCA が開発した「YMCA 英語能力検定テスト」(通称 YTEP)[3] がある。YTEP は他の能力テストには見られないクローズ形式と書き取りテストが中に含まれている。YTEP は専門学校の学生（高卒以上）全員と，会話科（大学生，一般社会人）の中級以上の学生に対して行なわれている。1985年度の受験者は 5,313人であった。

3.3 能力テストの選択

以上あげた能力テストが本稿の条件をみたしているかどうかをまとめたものが表―1 である。表が示すように，条件を満たしているのが YTEP である。また，YTEP の受験者は，大部分が専門学校で学ぶ学生であり，年代的にも本学学生と似たグループであるとみなせ，本学学生との能力差を比較する上で，興味深いものとなると考えた。以上の理由から YTEP を採用した。

3.4 方　法

1985年11月25日，27日，北星学園英文学科1年生 127 名に対して「YMCA 英語能力検定テスト」(YTEP) を実施した。時間は約70分，3クラスそれぞれ別々の時間を設定し，行った。試験は，テープレコーダー (Sony TC-1290) 1台を使用し，受験の際の注意，問題の指示などはあらかじめテープに録音されたもの（英語と日本語）を流した。

3.5 能力テストの形式

YTEP は次の4つの部分から構成されている。

パート 1. 文法 (Recognition test) 30問
パート 2. 読解力 (Reading comprehension test) 25問
パート 3. 聴解力 (Aural comprehension test) 55問
パート 4. 書き取り (Dictation) 6問

3.5.1 文法テストの形式

文法テストは中学，高校で学んだ英語の基本的な知識を問うもので，2つの形式からなっている。第一のタイプは，文があたえられ，その一部分が空白になっており，選択肢から適当な単語や句を選んで穴埋めをするもの。(付録1)

もう一つのタイプは，文の傍線部分の単語や句と同じ意味のものを選択肢から選ぶものである。(付録2)

3.5.2 読解力テストの形式

TOEFL などに見られる，パラグラフをあたえてその内容についての問が用意されている従来の形式とちがって，YTEP ではクローズ形式を読解力テストに導入している。パラグラフ導入部分のいくつかの文をそのままにし，ある部分から7語ごとに単語を空白にし，それを埋め，

表―1　各能力テストと本稿条件の関係

	TOEFL	TOEIC	英検(STEP)	JACET	YTEP
1. 信頼性・妥当性	○	○		○	○
2. クローズ・書き取り					○
3. 文法力等の個別テスト	○	○	○		○
4. データの処理				○	○
5. 学習効果を考慮しなくてよい				○	○
6. 受験料			○	○	○

パラグラフを完成させるのである。（付録３）

3.5.3　聴き取りテストの形式[4]

問題数も一番多く，YTEP の中で一番大きな比重をしめる聴き取りテストは，３つの形式からなっている。第１のタイプは，テープから流れてくる英文を聞いて適当な答を選択肢の中から選ぶもの。（付録４）

第２のタイプは会話形式で，テープを聞いた後，英文の問題を読みその答を選択肢から選ぶもの。（付録５）

３番目のタイプは，駅や空港などのアナウンスなど，英語が実際の場面で使われていることを想定したもので，状況をしっかりと認識した上で問に答えなければならない。（付録６）

3.5.4　書き取りテストの形式

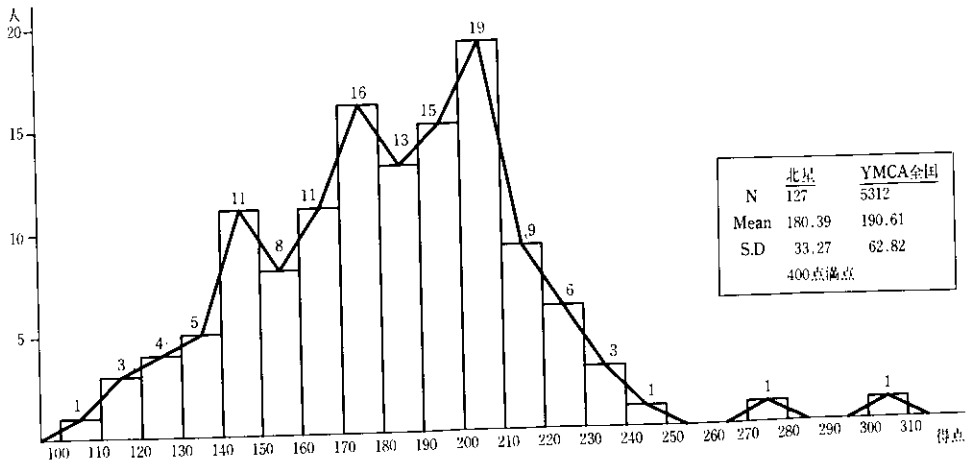
パート１～３まではマークシートに解答をする訳だが，書き取りテストでは，マークシートとは別のもう一枚の用紙に問題と解答欄が用意されている。用紙には，ひとかたまりの文章があたえられており，その中のいくつかの部分（文または句）が空白になっている。その空白の文，または句を書き取る訳だが，まずテープから文章全体が読まれ，内容を理解し，次に空白の部分が２度づつ読まれるのでそれらを書き取っていく。最後にまた文章全体が読まれ，書き直しをする時間があたえられる。（付録７）

4.　結果と考察

図－１は本学学生の得点分布をグラフにしたものである。図が示すように，205点の地点でピークが現われ，この値をはさんで左側は比較的なめらかなカーブをえがいて下降しゼロとなる。一方，ピークの右側は急激に下降し，いったんゼロとなり，さらに右すそに小さな山が２つ現われゼロとなる。平均が180.39点であるからグラフは非対称で，平均値よりも素点で25点高い所で最大となり，比較的よくできる学生がかたまっていることがわかる。

そして群をぬいて得点が高い学生が２人ほどいる。本学学生の受験者中，最高点を出した学生は，全国順位で 5,313 名中 305 位で　YMCA が設けている検定級[5] でもＡ級にランクされる。一方カーブの左側すそには，右で見られた様なはなれた山は見られない。つまり，他の学生と比較して極端に能力が落ちる学生はいないということで，入学時の選別が意図された機能をはたしているともいえる。

表－２は，各パートごとの平均点，標準偏差，最高，最低得点を本学学生と全国平均を比較して表にしたものである。平均の欄を見ると，４つのパートのうち全国平均を上回っているのはパート２の読解力のみである。このパートは読



図－１　YTEP における本学学生の得点分布

表—2 YTEP 集 計 表

	平 均		標準偏差		最高得点		最低得点	
	北 星	全 国	北 星	全 国	北 星	全 国	北 星	全 国
パート (60) 文 法	29.763	31.083	7.316	11.675	46	60	14	0
パート2 (50) 読解力	25.795	25.457	6.085	8.534	40	48	12	0
パート3 (180) 聴解力	76.102	84.546	19.472	28.812	129	174	30	0
パート4 (110) 書き取り	48.724	49.557	12.091	20.636	96	110	14	0
合 計 (400)	180.385	190.607	33.273	62.822	303	389	104	24

解力と呼ばれてはいるが，問題の形式は前述の様にクローズの形をとっており，Oller (1979) のいう grammer of expectancy (予測文法) を測定していると見ることができる。点数を見てみると，本学学生25.795点に対し，全国25.457点となっている。ただし，その差はわずか0.338点であるから，ほぼ同じと考えられる。

同様に，言語能力の中核となっている部分 (Oller 1979)，すなわち総合能力を測定しているとみられるパート4の書き取りテストを見ると，本学学生が48.724点に対し，全国49.557点で，パート2のクローズとは逆に，本学学生は0.833点全国平均より低い。ただし，その差はパート2同様わずかであるから，ほぼ同程度と見ることができよう。

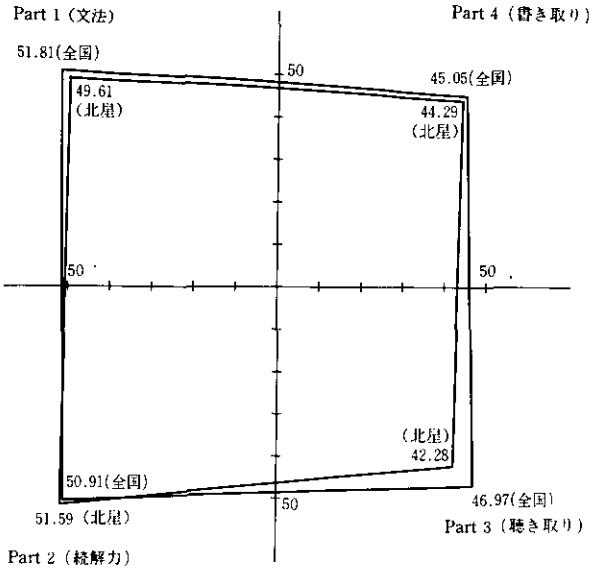
パート1の文法力テストでは，本学学生は全国平均より1.32点低い。しかしその差1.32点は文法力に大きな差があるとは考えにくい。

他のパートに比べて配点の高いパート3の聴解力テスト (180点満点) では，全国平均が84.546点に対し，本学学生は76.102点で8.444点全国を下回る。他のパートが全国平均と1点前後の差でほぼ同程度であったことを考えると聴解力に関して本学学生の力は，専門学校生を中心とした学生のそれより，明らかに低いと見てよいだろう。

ここで注意しなければならないことは，総合能力が同程度でありながら，聴解力のみが相対的に低いという事実である。今回の能力テストのみからでは，はっきりとした要因をつきとめるに足る十分なデータが得られなかったが，あえてその要因について推量してみるならば，次の様にも考えられる。本学学生の多くは，高等学校でほとんど聴解訓練を受けてこなかったため，入学時点から本学のカリキュラムでその訓練を始めても，4月から11月までの8カ月でその成果が十分に現われなかった。言いかえるならば，カリキュラムが8カ月で本学学生の聴解力を全国レベルまで引き上げるのに十分なものではない，と考えることもできる。いずれにしろ，クローズや書き取りが測定していると思われる英語総合能力がありながら，聴解力が相対的に低いということは，リスニングの訓練が足りないと言えるであろう。

表—2から各パートの平均得点率を出し，グラフにしたものが図—2である。各パート，バランスよく点数をとると正方形に近くなる。又，正方形が大きいほど能力が高いことを示し，その反対に小さな正方形は低い能力ということになる。4つのパートのうち，得点率が他の3つと比べて1つだけ低いという場合は，4辺形はいびつになる。

そこで図—2を見てみると，本学学生の4辺形は，全国平均のそれより一回り小さく，パート2の読解力でほんのわずかはみ出ている。つまり，本学学生の場合，聴解力が4辺形をいび



図—2 YTEP 各パート毎の平均得点率

つにしており，この図からもはっきりと，リスニングの力が相対的に弱いということが読みとれる。

5. む す び

クローズテスト，書き取りテストとも，専門学校の学生を中心とした全国平均と比較し，ほぼ同じ点数ということから考えて，本学学生の英語総合能力は，全国平均のそれとほぼ同程度ということができよう。しかしながら，聴き取りの力に関しては，専門学校を中心とした学生が本学学生を上回っていることが明らかになった。

限られたデータからあえてその要因をさぐってみるならば，本学カリキュラム上の英語運用面が，かならずしも十分でなかったと考えられ，したがってその部分の訓練が足りなかったためにおきた結果と思われる。英語との接触時間が専門学校と比較して少なく，特に聴解力を高める訓練となる Oral-Aural の授業数が少ない。また1クラスあたりの人数も本学の場合，平均して専門学校のそれに比べ多く，授業間のコーディネイションも専門学校に比べ弱いように思われる。この様なことが起因して聴解力の差となってあらわれたと見ることができよう。

しかしながら，本学学生間における能力の差は，極端なものではなく，ほぼ等質の学生が集まっており，入学時の選別が意図したように機能していると思われる。

今回の能力テストで明らかになった結果をふまえて今後のカリキュラムを考えてみるならば，Oral-Aural を中心とした時間を増やすことであり，できることなら1クラスの人数を少なくし，ネイティブによるオーラルの授業と L. L. などの授業との連けいを密にし，それが有機的に働き，その結果として運用能力が高まるよう質的量的に学生の英語接触時間を増やすことではなかろうか。それはとりもなおさず，学生の英語総合能力を高めることになり，バランスのとれた英語能力を養うことになる。

理想的な英語教育の目標が，4技能（聴く，話す，読む，書く）の運用能力をバランスよく高めることにあるものとすれば，これに一歩でも近づくためには，カリキュラムの改善が急務となるであろう。

注

1) Harris (1969: 130) は学習効果について次の様に述べている。
……Thus we may expect subjects who

are repeating a test (whether with the same or a parallel form) to score somewhat higher than they did the first time, even if their knowledge of the subject being tested has not itself increased. Test users must therefore make allowance for "practice effect" when evaluating scores on "progress" or "exit" test; slight improvements in such scores quite possibly represent improvement in test-taking skills, not increased competence in subject-matter skills.

2) Harris は適切なテストに関し次の様に述べている。
All good tests possess three qualities: validity, reliability, and practicality. That is to say, any test that we use must be appropreate in terms of our objectives, dependable in the evidence it provides, and applicable to our particular situation.

3) YMCA が1967～1982年にわたって開発した英語能力検定テストである。開発の中心になった人物は，Randolf H. Thrasher, Jr. で，Michigan Test of English Language Proficiency（ミシガンテスト）の著者の1人でもある。

4) YTEP 聴解テストの中に数問，関西に住んでいる受験生にはなじみが深いが，それ以外の者にとってはあまり聞きなれない地名や駅名がでてくる。関西の受験生にとっては有利とも見えるが，テスト全体の比重から考えれば，それにより他の者が大きく不利に働くとも考えにくい。しかし，全国的な規模で行う場合は問題として残るように思える。

5) YMCA の定める検定級において，A級で総合得点が700点以上の者は，第2次の面接試験を受験することができる。

S級	A級保持者で2次試験合格者
A〃	650点以上
B〃	649～550
C〃	549～450
D〃	449～350
E〃	349点以下

付　録

1. My uncle has been talking with his client —— two hours now.
   a) since
   b) from
   c) for
   d) by

2. John made up his mind to go there.
   a) wished
   b) remembered
   c) prepared
   d) decided

3. Nicholas Rizos was not a tourist. He was in Athens to work. He had arrived from America the day before on a Greek cargo ship. During his last year in high school, his uncle had invited ① to spend a year in ② and to help him in ③ garage. Nicholas accepted the invitation ④ he wanted to become a ⑤ ;……

①	②	③
a) her	a) school	a) urcle's
b) him	b) America	b) Greek
c) people	c) Greece	c) a
d) father	d) work	d) his

④	⑤
a) though	a) mechanic
b) because	b) tourist
c) which	c) worker
d) that	d) captain

4. Where's Sakurajima?（テープ）
   a) In Kumamoto.
   b) In Hokkaido.
   c) In Nagano.
   d) In Kagoshima.

5. "Has the 12:17 Hikari left yet?"（テープ）
   "No, but it's coming now, you'd better hurny. Track 17."（テープ）
   (設問) Where does this conversation most probably take place?
   a) In an airport.
   b) In a railroad station
   c) In a book store
   d) In a bakery shop

6. We will soon be arriving in Okayama. The local train for Kure and Hiroshima leaves track 7 in 5 minutes at 9 o'clock.（テープ）
   (設問) Where will the train arrive?
   a) Hiroshima　b) Okayama

c) Kure d) Onomichi

(設問) What track does the local train leave from?

a) No. 5 c) No. 9

b) No. 7 d) No. 11

7. In 1830, there was an immense forest only a few miles from what is now the great city of Cincinnati. The few settlers were restless folk ① for new frontiers, father westward. One of these settlers lived alone ② by the great forest ③ ; no one had ever known him to smile or to speak a needless word……

(回答) ① who had repeatedly abandoned the civilized communities (テープ)

② in a house of logs surrounded on all sides (テープ)

③ He himself seemed part of the gloomy silence of that forest (テープ)

参 考 文 献

浅野 博, 他 (編) (1979).『英語指導法ハンドブック 4<評価編>』大修館.

Lado, R (1961). Language Testing. Longman, Green, Co.

Harris, D (1969). Testing English as a Secord Language. McGraw. Hill.

垣田直巳, 他 (1985)『英語の評価論』大修館.

日本 YMCA イングリッシュセンター『YMCA 英語能力検定テスト 受験のしおり』

Oller, J (1971). "Dictation as a Device for Testing Foreign Language Skills," English Language Teaching, 25, 254~259.

——— (1973). "Cloze Tests of Second Language Proficiency and What They Measure," Language Learning Vol. 23 # 1.

——— and P. Irvine, P. Afai (1974). "Cloze, Dictation, and the Test of English as a Foreign Language," Language Learning Vol. 24 # 2.

——— (1979). Language Tests at School. Longman.

TOEIC 運営委員会 (編) (1981).『ビジネスマンの国際英語力テスト』朝日イブニングニュース社.

Stubbs, J & Tucker, G (1974). "The cloze test as a measure of English Proficiency," Modern Language Journal 58, 239-241.